**README**

Languages used: Python, Java

Libraries Used: nltk

Pickle

BeautifulSoup

PrettyTable

To install libraries: $pip install <libname>

Steps to execute the project-

Code to be executed task wise

PHASE1

1. Task1
2. a) Indexing: /IR-Project/Phase1/Task1/TaskA/Task-a/Indexer/main.py

b) BM25: /IR-Project/Phase1/Task1/TaskA/Task-b/BM25/BM25.py

Cosine Similarity : /IR-Project/Phase1/Task1/TaskA/Task-b/CosineSimilarity/main.py

TF-IDF: /IR-Project/Phase1/Task1/TaskA/Task-b/TFIDF/tf\_idf\_score.py

1. Lucene

Import folder HW\_4 to eclipse and run it.

1. Task2

BM25 with Query Expansion.

/IR-Project/Phase1/Task2/Expanded Query program/bm25WithQueryExpansion.py

1. Task3
2. BM25 with stopping: /IR-Project/Phase1/Task3/Task3A/bm25.py
3. BM25 with stemming: /IR-Project/Phase1/Task3/Task3B/bm25.py

PHASE2

Evaluation: /IR-Project/Phase2/evaluation.py

Snippet Generation

/IR-Project/Snippet Generation/Snippet\_Generation.py

Output Files:

PHASE1

Task1: /IR-Project/Phase1/Task1\_Output

Task2: /IR-Project/Phase1/Task2\_Output

Task3: /IR-Project/Phase1/Task3\_Output

PHASE2

/IR-Project/Phase1/Evaluation Output (Precison and Recall for 7 runs and MAP,MRR,P@K)

Snippet Generation

/IR-Project/Snippet Generation/ Snippet Generation output

Workbook

1. RetrievalSystemRanking : Top 100 documents with the ranks for all queries for 8 runs. Each run in an individual sheet.

/IR-Project/Spreadsheets/ RetrievalSystemRanking

1. Evaluations: Precison and Recall for all queries for 7 runs.

/IR-Project/Spreadsheets/ Evaluations

1. P@K: Precision at Rank = 5 and 20 for all queries for all runs.

/IR-Project/Spreadsheets/P@K

1. MAP\_and\_MRR: Mean Average Precision and Mean Reciprocal Rank for all retrieval models.

/IR-Project/Spreadsheets/MAP\_and\_MRR